

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-73081

(P2002-73081A)

(43)公開日 平成14年3月12日 (2002.3.12)

(51)Int.Cl.

G 10 L 15/24

G 06 F 3/16

G 10 L 15/14

15/18

15/00

識別記号

3 2 0

F I

G 06 F 3/16

H 04 M 1/00

1/725

G 10 L 3/00

テ-コ-ト (参考)

3 2 0 H 5 D 0 1 5

H 5 K 0 2 7

審査請求 未請求 請求項の数 5 OL (全 6 頁) 最終頁に続く

(21)出願番号

特願2000-256653(P2000-256653)

(71)出願人 000233169

株式会社日立超エル・エス・アイ・システムズ

東京都小平市上水本町5丁目22番1号

(22)出願日

平成12年8月28日 (2000.8.28)

(72)発明者 塚田 俊久

東京都小平市上水本町5丁目22番1号 日立超エル・エス・アイ・システムズ内

(72)発明者 北爪 吉明

東京都小平市上水本町5丁目22番1号 日立超エル・エス・アイ・システムズ内

(74)代理人 100081938

弁理士 徳若 光政

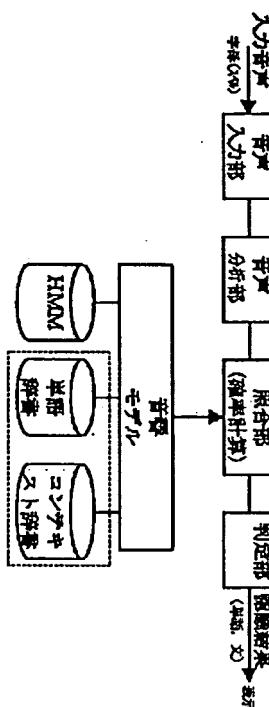
最終頁に続く

(54)【発明の名称】 音声認識方法と電子装置

(57)【要約】

【課題】 簡単な構成で英単語の識別率の大幅な改善を図った英語音声認識方法と音声認識技術を用いて使い勝手の改善を図った電子装置を提供する。

【解決手段】 英語音声認識方法において、英単語をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の发声の組み合わせに置き換えて入力するとともに、音声認識では上記英語のアルファベットの字母単位で行う。その音声認識方法を電子装置に搭載する。



(2)

【特許請求の範囲】

【請求項 1】 音声認識方法において、特定言語をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、音声認識では上記アルファベットの字母単位の組み合わせで行うようにしてなることを特徴とする音声認識方法。

【請求項 2】 請求項 1において、

上記複数言語での字母の入力は、英語のアルファベットの一部がギリシャ語のアルファベットに置き換えられて発声されるものであることを特徴とする音声認識方法。

【請求項 3】 請求項 1又は 2において、

上記音声信号の信号処理は、入力部においてデジタル信号化し、音声分析部において特徴抽出処理を行ない、照合部において予め用意された隠れマルコフモデルを含む音響モデルと照合することにより字母判定を行うものを含むことを特徴とする音声認識方法。

【請求項 4】 音声信号を取り込む音声入力部と、上記入力された音声信号をデジタル化し、その特徴抽出を行ない、予め用意された音響モデルと照合して字母判別を行う音声信号処理部を備え、特定言語の入力をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、上記音声認識部での音声認識では上記アルファベットの字母単位の組み合わせで行うようにしてなることを特徴とする電子装置。

【請求項 5】 請求項 4において、

上記複数言語での字母の入力は、英語のアルファベットの一部がギリシャ語のアルファベットに置き換えられて発声されるものであることを特徴とする電子装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音声認識方法と電子装置に関し、特に、音声入力を行うようにした携帯電話機等の電子装置に利用して有効な技術に関するものである。

【0002】

【従来の技術】音声認識技術は人が話す言葉や文章をコンピュータに直接認識させる技術であり、現在入力手段として用いられているキーボード入力、ペン入力などに替わる簡便な入力手段として注目されている。音声認識に関する文献の例として、Y. Obuchi, A. Koizumi, Y. Kitahara, J. Matsuda, and T. Tsukada, Proc. EUROSPEECH'99, pp. 2023-2026, 1999があり、口述ソフトウェアの例として、ViaVoice(IBM社)、NaturallySpeaking(Dragon社)、音声認識ソフトウェアの例としてASR 1600(L&H社)がある。上記IBM社の"ViaVoice"やDragon社の"Naturally Speaking"は主としてWintel PC向けのいわば重装備の口述ソフトウェア

2

である。

【0003】

【発明が解決しようとする課題】携帯電話機等のデジタル電子機器では、その小型化のためにパーソナルコンピュータのような多数のキーを持つキー・ボードを搭載することが難しいし、仮に多数のキーを実装できたとしても1つの大きさが小さく、かつ密集して配置されることなるために使い勝手の悪いものとなり現実的でない。そこで、入力手段として注目されている上記音声認識技術を用いることが考えられるが、上記の口述ソフトウェアは膨大な音声データを駆使することによりその認識率を高めたもので、コンテクスト(文脈)などの利用により文章入力についてはかなり高度の性能を有するものとなっているが、反面では大容量のメモリと高性能のCPUを必要とする。したがって、このような口述ソフトウェアを携帯電話機等のような小型で低消費電力であることが必要なデジタル電子機器に搭載することもやはり現実的ではない。

【0004】この発明の目的は、簡単な構成で英単語の識別率の大幅な改善を図った音声認識方法を提供することある。この発明の他の目的は、音声認識技術を用いて使い勝手の改善を図った電子装置を提供することある。この発明の前記ならびにそのほかの目的と新規な特徴は、本明細書の記述および添付図面から明らかになるであろう。

【0005】

【課題を解決するための手段】本願において開示される発明のうち代表的なものの概要を簡単に説明すれば、下記の通りである。音声認識方法において、特定言語をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、音声認識では上記アルファベットの字母単位の組み合わせで行う。

【0006】本願において開示される発明のうち他の代表的なものの概要を簡単に説明すれば、下記の通りである。入力部で音声信号を取り込んでデジタル化し、信号分析部で入力された音声信号の特徴抽出を行なって予め用意された音響モデルと照合して字母判別を行う電子装置において、特定言語の入力をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、上記音声認識部での音声認識では上記アルファベットの字母単位の組み合わせで行う。

【0007】

【発明の実施の形態】図1には、この発明に係る音声認識方法を説明するための一実施例のブロック図が示されている。この実施例の英語音声認識の方法は、図1の各ブロックでの信号処理に沿って行われる。発聲音はまず

(3)

3

音声入力部においてデジタル信号化される。この実施例では、簡単な構成での高い音声識別率を実現するため、音声入力を特定言語の単語等ではなく、そのスペルに対応したアルファベットの字母の単位で行うようにすることに1つの特徴としている。

【0008】ここで、「字母」とは、国語辞典によれば、「かな」、「アルファベット」、「梵字」等のように発音を示すつづり字のひとつひとつのことをいい、英語では「phoneme」(フォニーム)に相当する。このフォニーム(phoneme)は、音素のことをいい、ある言語の音声学上の最小単位を意味するものである。

【0009】上記のような字母単位でのスペル音声入力には、字母毎に一定の無音期間を挿入するか、あるいは字母の区切りを意味するキー入力信号を挿入すること等により行われる。使い勝手を考慮すれば、字母毎に無音期間を挿入することが有益であると考えられるが、より確実な字母の区切りを行うなら、キー信号を用いることが有益である。

【0010】上記音声入力部から字母単位で入力された音声信号は、音声分析部に送られ、そこで特徴抽出処理が行われる。具体的には短時間周波数分析である。分析結果は照合部において予め用意された音響モデルと照合され、判定処理を行う。最も高いスコアを得たものが認識結果として表示される。

【0011】照合部において用いられる音響モデルは、一般的にはHMMと単語辞書とコンテキスト辞書を連結したものである。ここで、HMMとは、隠れマルコフモデル(Hidden Markov Model)と呼ばれるもので認識の基となる参照モデルである。このHMMに単語辞書や文脈データなどを組んだコンテキスト辞書を組み合わせたものが音響モデルとなる。

【0012】上記音声認識の基本となるHMMは比較的軽いソフトウェアであるが、従来のように口述に対応した特定言語の音声識別を行うようにするには、その言語の単語辞書やコンテキスト辞書を設けることが必須となり、そのアプリケーションによっては重くなりやすい。長文読み上げを実時間で認識することが求められる口述ソフトウェアなどでは、単語辞書は無論のこと文脈や文例などのデータを大量に設けることが必要になるため、データを高速で処理する高性能のCPUおよび大容量のメモリが必要になる。

【0013】携帯型電子装置においては長時間動作を可能にするため、部品点数を減らし消費電力を押さえる必要がある。メモリの容量はできるだけ少なくし、CPUのパワーも制限される。すなわち辞書を始めとするデータ量はなるべく軽いものが求められる。そこで、この発明に係る英語音声認識方法では、英単語を字母単位で入力し、それを字母単位で音声識別を行うようにすることにより、基本的には同図で点線で示したように単語辞書やコンテスト辞書を大幅に簡略化するものである。

4

【0014】この実施例では、英語の音声による入力能率を極限まで高めるためにスペル発声による音声入力が採用される。例えば“butter”は“bi: ju: ti: ti: i: a:r”と発声する。これは単に“b*t*r”(バタ-)と発声するよりも情報量が多いためにその認識率は格段に改善される。発声するのはアルファベットの26文字だけなので、あらゆる単語の発音記号とそれに付随した音声データを駆使する重装備のソフトウェアを必要とするものに比較して、はるかに軽量のソフトウェアであらゆる単語を入力することができかつ正確に認識される。

【0015】単語の入力が正確にできれば文章の入力も同様に入力できる。これはいわばワードプロセッサの音声版とみなすことができる。英語のワードプロセッサが行っていることはまさにこのことであって“butter”を入力するにあたっては“b” “u” “t” “i” “e” “r”とキーボードのキーを叩くのである。

【0016】本発明のもうひとつの特徴は、上記に加えてギリシャ語のアルファベットを併用することである。先の“butter”的いうと“b”、“u”、“t”、“e”、

“r”等の代わりに、 β (beita:ベータ)、 ν (ju:psilan:イフ・シラ)、 τ (tau:タウ)、 ϵ (eps*lan:エプシロン)、 ρ (rou:ロ)等を併用することである。“butter”を“bi: ju: t i: ti: i: a:r”と発声したときその出力が“gutter”となつたとする。つまり、“b”と“g”が混線したわけである。このとき続けて“ β ”と発声すれば出力は自動的に正しい“butter”となる。このような混線が予め想されるときには最初から“beita ju: ti: ti: i: a:r”と発声することで混線をなくすることもできる。 β (b)だけでなく、 δ (d)、 γ (g)、 π (p)、 τ (t)等も利用する。この他にも α (a)、 κ (k)、 ν (n)、 μ (m)、 ω (o)、 ρ (r)、 σ (s)等も適宜利用することが可能である。

【0017】上記のような認識方法を探ることにより、英語などの単語認識率を極限にまで高めることができる。すなわちスペリングによる入力においては認識すべき音節が基本的にはアルファベットに限定されるため認識率が高まる。さらに、単語に該当しないアルファベットの組み合わせが除外されることも認識率を高めるのに効果がある。例えば、“butter”を“bi: ju: ti: ti: i: a:r”と発声したとき、最初の“b”を通常混同しやすい“d”と間違えることはない。なぜなら“dutter”という単語がないからである。

【0018】スペリング入力はそれだけでも効果があるが、さらにこれを効果的にするためにギリシャ語のアルファベットを併用するものである。これはアルファベット間の違いをさらに高めるのに効果的である。英語を例にとればいわゆる「e問題」を避けることができる。すなわち、“b”、“d”、“e”、“g”、“p”、“t”等のまぎらわしさを“ β ”、“ δ ”、“ ϵ ”、“ γ ”、“ π ”、“ τ ”等の併用により避けることができる。英語とギリ

(4)

5

シャ語のアルファベットの中からお互いに距離の離れた
すなわち類似性の少ないものを選ぶことができるからで
ある。ギリシャ語のアルファベットは英語ほどには知ら
れていないが、それでも比較的なじみがあり記憶しやす
い。

【0019】なお、本発明は英語に限らず、仏語、独
語、ロシア語等の欧米系言語に適用可能であるという普
遍性を持つ。すなわちたとえば仏語とギリシャ語のアル
ファベットを組み合わせて使うことにより、仏語の単語
認識率を究極まで高めることができる。

【0020】本発明に係る音声認識方法に対応したソフ
トウェア等を携帯型電子装置に搭載した場合、単語の認
識率が改善されることによる音声入力の能率向上効果が
大きいことは無論であるが、それ以上に大きいのは必ず
入力できるという安心感である。これは製品を使用する
立場からすると非常に大きなことで、何度も発声しても正
しい入力ができないのでは使ってもらえない。このことは本発明の適用製品が何であっても言えることである。

【0021】本発明に関わる音声認識方法では、前記の
ようなスペリング入力にいくつかのコマンド、例えば
“Capital letter”（大文字）、“hyphen”（ハイフ
ン）、“comma”（コンマ）、“period”（ピリオド）、
“colon”（コロン）、“space”（スペース）、“new pa
ragraph”（改行）等を付加するようにしてもよい。この
ようなコマンドを設けることにより、文章入力も容易に
できる訳でありその効果は極めて大きい。更に従来の手
段では入力すべき単語のデータを予め入力する必要があ
り、その認識率は単語の数が1000語、10000語と増えるにしたがって顕著に減少していった。本発明を用いれば単語数が増大しても認識率は変化せず必ず入力
できる。

【0022】この実施例の音声認識方法では、基本的に
は英語やギリシャ語のアルファベット等のように少ない
数の音声識別を基本としているので、音響モデルのデータ
量を極力少なくできる上に、かかるアルファベットの字母
結果を組み合わせることで、結果的にあらゆる種類
の単語や文章も入力することができる。このように音声
認識でのデータ量を少なくすることができるので、それを
処理する中央処理装置CPUも低消費電力のRISC
(Reduced instruction set computer)タイプのものを用
いることができ、しかもメモリ容量も少なくてよい。この
結果、この発明に係る音声認識方法は、携帯用電子装置
に最適な入力方法である。使い勝手を良くするために、
特定の制御信号や動作命令を音声で行うようにした
場合でも、単語辞書やコンテキスト辞書は小規模で済
む。

【0023】図2には、この発明を携帯型通訳機に適用
した場合の一実施例の外観図が示されている。この実施
例の携帯型通訳機は、英語-日本語通訳に向けられて
いる。この実施例の携帯型通訳機の使用方法は次の通りで

6

ある。発声鉤1を押してマイク2に向かって発声する
と、認識結果が表示装置3に表示される。正しい結果が
得られたらOK鉤4を押して文例検索に移行する。スク
ロール鉤5により検索し所望の文が見つかるとOK鉤4
を押して訳文表示をする。さらに発声鉤1を押すと訳文
の音声がスピーカ6を通して流れる。

【0024】この携帯型通訳機に本発明に係る音声認識
方法を適用した例を“Thank”を例にして示す。発声鉤1
を押してマイク2に口を近づけて“ti: eit* ei en ke
10 i”（ティ-エイチ、アイ、イ-、エイ、ケイ）と発声する。認識結果
が“Thank”と表示装置3に表示される。認識時には自動
的にスペルチェックが行われ、辞書にない単語は除外さ
れる。これは認識率を高めるのに効果がある。認識結果
を表示装置3でチェックしたらOK鉤4を押すと文例が
表示される。

【0025】“Thank you for your help.”を選んでOK
鉤4を押すと「ありがとう。助かりました。」と訳文
が表示される。そこで発声鉤1を押すと音声がスピーカ
6から流れる。文例検索による選択例について述べたが
20 短い文であれば“ti: eit* ei en kei, space, wai ou
ju:”といった具合に入力することもできる。表示装置3
には“Thank you”と表示される。音声出力は「ありがとう」
となる。ここで、アルファベットhに対応した発音
eit*において、発音記号の部分が*で置き換えている。
このことは、前記のギリシャ語アルファベットの発音記
号も同様である。以下同じ。

【0026】基本的にはスペル入力で入力はできるが、
周囲の雑音などの関係で入力しづらい場合もある。この
ようなときにはギリシャ語のアルファベットを併用する
30 のが有効である。先の“thank”を例にとれば“th”、
“a”、“n”、“k”的代わりにθ「しーた」、α「あるふ
あ」、ν「にゅー」、κ「かっぱ」を用いる、等である。
また、修正時に用いるのも有効である。同じ例で結果
の表示が“think”になった時に“a”的部分をαに置き
換えて入力し直すことにより正しい結果“thank”が容易
に得られる。

【0027】英語のアルファベットに対応するギリシャ
語のアルファベットが必ずしもある訳ではないので併用
が基本となる。英語とギリシャ語のアルファベットの組
40 み合わせにより「e問題」を避けることができるばかり
でなく、アルファベット間の距離を増大することができ
結果として認識率が究極まで高められる。ここで、「e
問題」とは、b、d、e、g、p、t等の発声の類似性
が認識率を妨げるという問題のことである。

【0028】前記図2の携帯型通訳機をそのまま携帯型
のワードプロセッサとして用いるようにすることもできる。
その使用方法は、次の通りである。まずコマンドとして
“Word Processor”と発声し、続いて“File”そして
“New”と発声する。新規入力画面が表示される。“Ti
50 tle of the invention”と発声する。その結果が“Idol

(5)

7

of a convention" と表示されたとする。スクロール鉤によりポインターを "Idol" の "I" に合わせ、 "ti: ai ti: el i:" (タイ アイ テイ エル イ) もしくは "tau iota tau lambda epsilon" (タウ イオタ タウ ラムダ エプシロン) 等と発声する。

【0029】あるいは、英語とギリシャ語のアルファベットの組み合わせ発声でもいい。 "tau ai tau el i:" (タウ アイ タウ エル イ) となる。これにより "Title of a convention" が表示される。次にポインターを "a" に合わせ "θ eit* i:" (シク エイ イ) 等と発声する。 "a" が "the" に代わる。以下同様にして正しい入力 "Title of the invention" が得られる。続いて本文の入力に入り、文章入力、適宜修正過程を経て入力を完成する。途中あるいは最初からスペル入力を用いることもある。辞書にはない技術用語や人名、地名の入力時等にはとくに便利である。全文が完成したら "Store" (ストア) と発声し、入力した文面を保存し作業を終了する。

【0030】このような携帯型ワードプロセッサでは、文章読み上げ入力とスペル入力を併用しているので、入力効率が改善されるという効果がある。また、スペル入力を随時採用すれば周囲を気にすることなく入力作業を行うことができる。ギリシャ語のアルファベットを頻繁に使えばこの効果を更に高めることができる。

【0031】上記図2の携帯型通訳機は、そのまま携帯電話機能もったパームトップ型パーソナルコンピュータに置き換えることができる。例えば、電子メール (e-mail) テキスト文の入力に本発明を適用した例をもうひとつの使用方法を以下に説明する。

【0032】まずe-mailのコマンドとして "e-mail" と発声する。画面がメール新規作成に切り替わる。表示画面に表示されたアドレス帖からアドレスを選択する。本文の入力に入る。 "How are you" "Question mark" と発声する。 "How are you?" と表記される。もしも、修正する必要があるときは随時本発明に係るグレコロマン式スペル入力により修正する。以下同様にしてメールの文章を入力する。最後にコマンドとして "Send mail" と発声するとメールが送信される。

【0033】上記のパームトップ型パーソナルコンピュータは、携帯電話機であってもいいし、ネットワーク接続されたノート型パーソナルコンピュータ等であってもいい。また、本発明を用いる副次的效果としては入力をスペル方式に適宜変更することにより、側で聞いている人がいても気にすることなく使用することができる。

【0034】上記の実施例から得られる作用効果は、下記の通りである。

(1) 音声認識方法において、特定言語をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、音声認識では上記アルファベットの字母単位の組み合わせで行うこと

8

により、簡単な構成で識別率の大幅な改善を図ることができるという効果が得られる。

【0035】(2) 上記に加えて、上記複数言語での字母の入力は、英語のアルファベットの一部がギリシャ語のアルファベットに置き換えることによる類似性の高い字母の識別が簡単となり、1回の認識での識別率の大幅な改善につながるという効果が得られる。

【0036】(3) 上記に加えて、上記音声信号の信号処理は入力部においてデジタル信号化し、音声分析部において特徴抽出処理を行ない、照合部において予め用意された隠れマルコフモデルを含む音響モデルと照合することにより簡単な信号処理での認識が可能となり、メモリやCPUに安価なものを用い、簡易なソフトウェアでの認識が可能になるという効果が得られる。

【0037】(4) 音声信号を取り込んでデジタル化する音声入力部と、上記デジタル化された音声信号の特徴抽出を行ない、予め用意された音響モデルと照合して字母判別を行う音声信号処理部を備え、特定言語の入力をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、上記音声認識部での音声認識では上記アルファベットの字母単位の組み合わせで行うことにより、キー操作が簡単で文字や単語入力を簡単に行える電子装置を得ることができるという効果が得られる。

【0038】(5) 上記に加えて、上記複数言語での字母の入力は、英語のアルファベットの一部がギリシャ語のアルファベットに置き換えられて発声することにより、簡単な信号処理により識別率の大幅な改善につながるという効果が得られる。

【0039】以上本発明者よりなされた発明を実施例に基づき具体的に説明したが、本願発明は前記実施例に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。例えば、前記のような英語、仏語、独語、ロシア語等の欧米系言語に適用可能であるという普遍性を持つものである。すなわち、たとえば仏語とギリシャ語のアルファベットを組み合わせて使うことにより、仏語の単語認識率を究極まで高めることができる。この発明に係る音声認識方法を用いた音声認識機能が搭載される電子装置は、前記の実施例の他にカーナビゲーションシステム等にも同様に適用できる。

【0040】

【発明の効果】本願において開示される発明のうち代表的なものによって得られる効果を簡単に説明すれば、下記の通りである。音声認識方法において、特定言語をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、音声認識では上記アルファベットの字母単位の組み合わせで行

(6)

9 うことにより、簡単な構成で識別率の大幅な改善を図ることができる。

【0041】音声信号を取り込んでデジタル化する音声入力部と、上記デジタル化された音声信号の特徴抽出を行ない、予め用意された音響モデルと照合して字母判別を行う音声信号処理部を備え、特定言語の入力をそれを構成するアルファベットに置き換えて音声入力するとともに、かかるアルファベットを複数言語の字母の発声の組み合わせに置き換えて入力するとともに、上記音声認識部での音声認識では上記アルファベットの字母単位の組み合わせで行うようにすることにより、キー操作が簡

10

單で文字や単語入力を簡単に行える電子装置を得ることができる。

【図面の簡単な説明】

【図1】この発明に係る音声認識方法を説明するための一実施例を示すブロック図である。

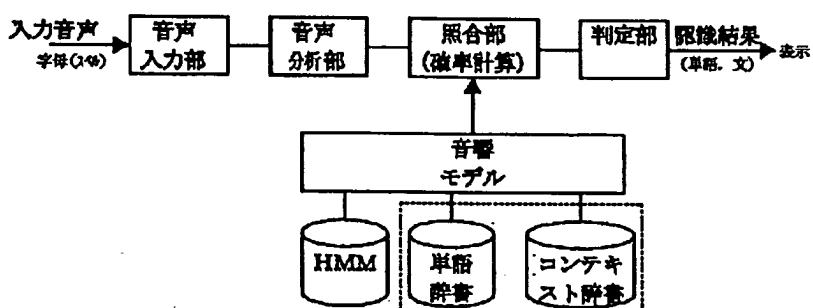
【図2】この発明を携帯型通訳機に適用した場合の一実施例を示す外観図である。

【符号の説明】

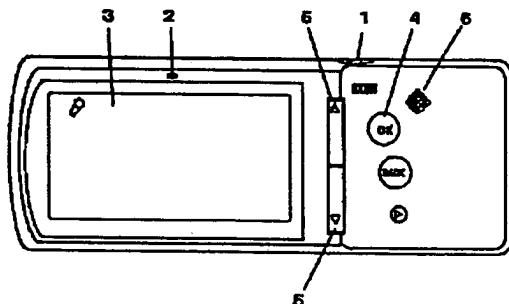
1…发声鉤、2…表示装置、3…表示画面、4…OK

10 鉤、5…スクロール鉤、6…スピーカー。

【図1】



【図2】



フロントページの続き

(51) Int. Cl. 7

H 04 M 1/00
1/725

識別記号

F I

G 10 L 3/00

テマコード* (参考)

537H
551A
551B

(72) 発明者 田中 誠

東京都小平市上水本町5丁目22番1号 日
立超エル・エス・アイ・システムズ内

(72) 発明者 内館 秀樹

東京都小平市上水本町5丁目22番1号 日
立超エル・エス・アイ・システムズ内

F ターム(参考) 5D015 AA05 BB02 HH23 KK02
5K027 AA11 BB01 HH20